RAW SEQUENCE LISTING

The Biotechnology Systems Branch of the Scientific and Technical Information Center (STIC) no errors detected.

Application Serial Number: _____ 10/572,638 _____
Source: _____ IFWP _____
Date Processed by STIC: _____ 02/10/2007 _____

ENTERED

RAW SEQUENCE LISTING   DATE: 02/10/2007
PATENT APPLICATION: US/10/572,638   TIME: 09:56:30

Input Set : N:\DA\pto_da.txt
Output Set: N:\CRF4\02102007\J572638.raw

```
 3 <110> APPLICANT: DUKE UNIVERSITY
 5 <120> TITLE OF INVENTION: CONSENSUS/ANCESTRAL IMMUNOGENS
 7 <130> FILE REFERENCE: 01579-1093
 9 <140> CURRENT APPLICATION NUMBER: 10/572,638
C--> 10 <141> CURRENT FILING DATE: 2006-03-17
12 <150> PRIOR APPLICATION NUMBER: PCT/US2004/03097
13 <151> PRIOR FILING DATE: 2004-09-17
15 <150> PRIOR APPLICATION NUMBER: 60/604,722
16 <151> PRIOR FILING DATE: 2004-08-27
18 <150> PRIOR APPLICATION NUMBER: 60/503,460
19 <151> PRIOR FILING DATE: 2003-09-17
21 <160> NUMBER OF SEQ ID NOS: 321
23 <170> SOFTWARE: PatentIn Ver. 3.3
25 <210> SEQ ID NO: 1
26 <211> LENGTH: 866
27 <212> TYPE: PRT
28 <213> ORGANISM: Human immunodeficiency virus
30 <400> SEQUENCE: 1
31 Met Arg Val Met Gly Ile Gln Arg Asn Cys Gln His Leu Trp Arg Trp
32  1               5                   10                  15
34 Gly Thr Met Ile Leu Gly Met Leu Met Ile Cys Ser Ala Ala Glu Asn
35               20                  25                  30
37 Leu Trp Val Thr Val Tyr Tyr Gly Val Pro Val Trp Lys Glu Ala Asn
38           35                  40                  45
40 Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys Ala Tyr Asp Thr Glu Val
41       50                  55                  60
43 His Asn Val Trp Ala Thr His Ala Cys Val Pro Thr Asp Pro Asn Pro
44  65                  70                  75                  80
46 Gln Glu Ile Val Leu Glu Asn Val Thr Glu Asn Phe Asn Met Trp Lys
47               85                  90                  95
49 Asn Asn Met Val Glu Gln Met His Glu Asp Ile Ile Ser Leu Trp Asp
50              100                 105                 110
52 Gln Ser Leu Lys Pro Cys Val Lys Leu Thr Pro Leu Cys Val Thr Leu
53          115                 120                 125
55 Asn Cys Thr Asn Val Arg Asn Val Ser Ser Asn Gly Thr Glu Thr Asp
56       130                 135                 140
58 Asn Glu Glu Ile Lys Asn Cys Ser Phe Asn Ile Thr Thr Glu Leu Arg
59 145                 150                 155                 160
61 Asp Lys Lys Gln Lys Val Tyr Ala Leu Phe Tyr Arg Leu Asp Val Val
62              165                 170                 175
64 Pro Ile Asp Asp Lys Asn Ser Ser Glu Ile Ser Gly Lys Asn Ser Ser
65          180                 185                 190
67 Glu Tyr Tyr Arg Leu Ile Asn Cys Asn Thr Ser Ala Ile Thr Gln Ala
```

RAW SEQUENCE LISTING   DATE: 02/10/2007
PATENT APPLICATION: US/10/572,638   TIME: 09:56:30

Input Set : **N:\DA\pto.da.txt**
Output Set: **N:\CRF4\02102007\J572638.raw**

```
68        195                200                205
70 Cys Pro Lys Val Ser Phe Glu Pro Ile Pro Ile His Tyr Cys Ala Pro
71     210                215                220
73 Ala Gly Phe Ala Ile Leu Lys Cys Asn Asp Lys Lys Phe Asn Gly Thr
74 225                230                235                240
76 Gly Pro Cys Lys Asn Val Ser Thr Val Gln Cys Thr His Gly Ile Lys
77           245                250                255
79 Pro Val Val Ser Thr Gln Leu Leu Leu Asn Gly Ser Leu Ala Glu Glu
80           260                265                270
82 Glu Ile Ile Ile Arg Ser Glu Asn Ile Thr Asn Asn Ala Lys Thr Ile
83     275                280                285
85 Ile Val Gln Leu Asn Glu Ser Val Glu Ile Asn Cys Thr Arg Pro Asn
86     290                295                300
88 Asn Asn Thr Arg Lys Ser Ile His Ile Gly Pro Gly Gln Ala Phe Tyr
89 305                310                315                320
91 Ala Thr Gly Glu Ile Ile Gly Asp Ile Arg Gln Ala His Cys Asn Ile
92           325                330                335
94 Ser Arg Thr Lys Trp Asn Lys Thr Leu Gln Gln Val Ala Lys Lys Leu
95           340                345                350
97 Arg Glu His Phe Asn Asn Lys Thr Ile Ile Phe Lys Pro Ser Ser Gly
98     355                360                365
100 Gly Asp Leu Glu Ile Thr Thr His Ser Phe Asn Cys Gly Gly Glu Phe
101    370                375                380
103 Phe Tyr Cys Asn Thr Ser Gly Leu Phe Asn Ser Thr Trp Met Phe Asn
104 385                390                395                400
106 Gly Thr Tyr Met Phe Asn Gly Thr Lys Asp Asn Ser Glu Thr Ile Thr
107           405                410                415
109 Leu Pro Cys Arg Ile Lys Gln Ile Ile Asn Met Trp Gln Gly Val Gly
110           420                425                430
112 Gln Ala Met Tyr Ala Pro Pro Ile Glu Gly Lys Ile Thr Cys Lys Ser
113    435                440                445
115 Asn Ile Thr Gly Leu Leu Leu Ile Thr Arg Asp Gly Gly Asn Asn Ser
116    450                455                460
118 Asn Lys Asn Lys Thr Glu Thr Phe Arg Pro Gly Gly Gly Asp Met Arg
119 465                470                475                480
121 Asp Asn Trp Arg Ser Glu Leu Tyr Lys Tyr Lys Val Val Lys Ile Glu
122           485                490                495
124 Pro Leu Gly Val Ala Pro Thr Lys Ala Lys Arg Arg Val Val Glu Arg
125           500                505                510
127 Glu Lys Arg Ala Val Gly Ile Gly Ala Val Phe Leu Gly Phe Leu Gly
128    515                520                525
130 Ala Ala Gly Ser Thr Met Gly Ala Ala Ser Ile Thr Leu Thr Val Gln
131    530                535                540
133 Ala Arg Gln Leu Leu Ser Gly Ile Val Gln Gln Ser Asn Leu Leu
134 545                550                555                560
136 Arg Ala Ile Glu Ala Gln Gln His Leu Leu Gln Leu Thr Val Trp Gly
137           565                570                575
139 Ile Lys Gln Leu Gln Ala Arg Val Leu Ala Val Glu Arg Tyr Leu Lys
140    580                585                590
```

```
142 Asp Gln Gln Leu Leu Gly Ile Trp Gly Cys Ser Gly Lys Leu Ile Cys
143         595             600                 605
145 Thr Thr Asn Val Pro Trp Asn Ser Ser Trp Ser Asn Lys Ser Gln Asp
146         610             615                 620
148 Glu Ile Trp Asp Asn Met Thr Trp Met Glu Trp Glu Arg Glu Ile Ser
149 625                 630             635                 640
151 Asn Tyr Thr Asp Ile Ile Tyr Arg Leu Ile Glu Glu Ser Gln Asn Gln
152             645             650                 655
154 Gln Glu Lys Asn Glu Gln Glu Leu Leu Ala Leu Asp Lys Ala Ser Leu
155         660             665                 670
157 Trp Asn Trp Phe Asp Ile Thr Asn Trp Leu Trp Tyr Ile Lys Ile Phe
158         675             680                 685
160 Ile Met Ile Val Gly Gly Leu Ile Gly Leu Arg Ile Val Phe Ala Val
161         690             695                 700
163 Leu Ser Ile Val Asn Arg Val Arg Gln Gly Tyr Ser Pro Leu Ser Phe
164 705             710                 715                 720
166 Gln Thr Leu Ile Pro Asn Pro Arg Gly Pro Asp Arg Pro Glu Gly Ile
167             725             730                 735
169 Glu Glu Glu Gly Gly Glu Gln Gly Arg Asp Arg Ser Ile Arg Leu Val
170             740             745                 750
172 Asn Gly Phe Leu Ala Leu Ala Trp Asp Asp Leu Arg Ser Leu Cys Leu
173         755             760                 765
175 Phe Ser Tyr His Arg Leu Arg Asp Phe Ile Leu Ile Ala Ala Arg Thr
176         770             775                 780
178 Val Glu Leu Leu Gly Arg Arg Ser Leu Arg Gly Leu Gln Lys Gly Trp
179 785             790                 795                 800
181 Glu Ala Leu Lys Tyr Leu Gly Asn Leu Leu Gln Tyr Trp Gly Gln Glu
182             805             810                 815
184 Leu Lys Asn Ser Ala Ile Ser Leu Leu Asp Thr Thr Ala Ile Ala Val
185         820             825                 830
187 Ala Glu Gly Thr Asp Arg Val Ile Glu Ile Val Gln Arg Ala Cys Arg
188         835             840                 845
190 Ala Ile Leu Asn Ile Pro Arg Arg Ile Arg Gln Gly Leu Glu Arg Ala
191     850             855                 860
193 Leu Leu
194 865
197 <210> SEQ ID NO: 2
198 <211> LENGTH: 2607
199 <212> TYPE: DNA
200 <213> ORGANISM: Human immunodeficiency virus
202 <400> SEQUENCE: 2
203 gccaccatgc gcgtgatggg catccagcgc aactgccagc acctgtggcg ctggggcacc 60
204 atgatcctgg gcatgctgat gatctgctcc gccgccgaga acctgtgggt gaccgtgtac 120
205 tacggcgtgc ccgtgtggaa ggaggccaac accaccctgt tctgcgcctc cgacgccaag 180
206 gcctacgaca ccgaggtgca caacgtgtgg gccacccacg cctgcgtgcc caccgacccc 240
207 aacccccagg agatcgtgct ggagaacgtg accgagaact tcaacatgtg gaagaacaac 300
208 atggtggagc agatgcacga ggacatcatc tccctgtggg accagtccct gaagccctgc 360
209 gtgaagctga cccccctgtg cgtgaccctg aactgcacca acgtgcgcaa cgtgtcctcc 420
210 aacggcaccg agaccgacaa cgaggagatc aagaactgct ccttcaacat caccaccgag 480
```

```
211 ctgcgcgaca agaagcagaa ggtgtacgcc ctgttctacc gcctggacgt ggtgcccatc 540
212 gacgacaaga actcctccga gatctccggc aagaactcct ccgagtacta ccgcctgatc 600
213 aactgcaaca cctccgccat cacccaggcc tgccccaagg tgtccttcga gcccatcccc 660
214 atccactact gcgcccccgc cggcttcgcc atcctgaagt gcaacgacaa gaagttcaac 720
215 ggcaccggcc cctgcaagaa cgtgtccacc gtgcagtgca cccacggcat caagcccgtg 780
216 gtgtccaccc agctgctgct gaacggctcc ctggccgagg aggagatcat catccgctcc 840
217 gagaacatca ccaacaacgc caagaccatc atcgtgcagc tgaacgagtc cgtggagatc 900
218 aactgcaccc gccccaacaa caacacccgc aagtccatcc acatcggccc cggccaggcc 960
219 ttctacgcca ccggcgagat catcggcgac atccgccagg cccactgcaa catctcccgc 1020
220 accaagtgga acaagaccct gcagcaggtg gccaagaagc tgcgcgagca cttcaacaac 1080
221 aagaccatca tcttcaagcc ctcctccggc ggcgacctgg agatcaccac ccactccttc 1140
222 aactgcggcg gcgagttctt ctactgcaac acctccggcc tgttcaactc cacctggatg 1200
223 ttcaacggca cctacatgtt caacggcacc aaggacaact ccgagaccat caccctgccc 1260
224 tgccgcatca agcagatcat caacatgtgg cagggcgtgg gccaggccat gtacgccccc 1320
225 cccatcgagg gcaagatcac ctgcaagtcc aacatcaccg gcctgctgct gacccgcgac 1380
226 ggcggcaaca actccaacaa gaacaagacc gagaccttcc gccccggcgg cggcgacatg 1440
227 cgcgacaact ggcgctccga gctgtacaag tacaaggtgg tgaagatcga gcccctgggc 1500
228 gtggccccca ccaaggccaa gcgccgcgtg gtggagcgcg agaagcgcgc cgtgggcatc 1560
229 ggcgccgtgt tcctgggctt cctgggcgcc gccggctcca ccatgggcgc cgcctccatc 1620
230 accctgaccg tgcaggcccg ccagctgctg tccggcatcg tgcagcagca gtccaacctg 1680
231 ctgcgcgcca tcgaggccca gcagcacctg ctgcagctga ccgtgtgggg catcaagcag 1740
232 ctgcaggccc gcgtgctggc cgtggagcgc tacctgaagg accagcagct gctgggcatc 1800
233 tggggctgct ccggcaagct gatctgcacc accaacgtgc cctggaactc ctcctggtcc 1860
234 aacaagtccc aggacgagat ctgggacaac atgacctgga tggagtggga gcgcgagatc 1920
235 tccaactaca ccgacatcat ctaccgcctg atcgaggagt cccagaacca gcaggagaag 1980
236 aacgagcagg agctgctggc cctggacaag tgggcctccc tgtggaactg gttcgacatc 2040
237 accaactggc tgtggtacat caagatcttc atcatgatcg tgggcggcct gatcggcctg 2100
238 cgcatcgtgt tcgccgtgct gtccatcgtg aaccgcgtgc gccagggcta ctcccccctg 2160
239 tccttccaga ccctgatccc caacccccgc ggccccgacc gccccgaggg catcgaggag 2220
240 gagggcggcg agcagggccg cgaccgctcc atccgcctgg tgaacggctt cctggccctg 2280
241 gcctgggacg acctgcgctc cctgtgcctg ttctcctacc accgcctgcg cgacttcatc 2340
242 ctgatcgccg cccgcaccgt ggagctgctg ggccgccgct ccctgcgcgg cctgcagaag 2400
243 ggctgggagg ccctgaagta cctgggcaac ctgctgcagt actggggcca ggagctgaag 2460
244 aactccgcca tctccctgct ggacaccacc gccatcgccg tggccgaggg caccgaccgc 2520
245 gtgatcgaga tcgtgcagcg cgcctgccgc gccatcctga acatcccccg ccgcatccgc 2580
246 cagggcctgg agcgcgccct gctgtaa                                     2607
249 <210> SEQ ID NO: 3
250 <211> LENGTH: 2535
251 <212> TYPE: DNA
252 <213> ORGANISM: Human immunodeficiency virus
254 <400> SEQUENCE: 3
255 gccgccatgc gcgtgatggg catcctgcgc aactgccagc agtggtggat ctggggcatc 60
256 ctgggcttct ggatgctgat gatctgctcc gtggtgggca acctgtgggt gaccgtgtac 120
257 tacggcgtgc ccgtgtggaa ggaggccaag accaccctgt tctgcgcctc cgacgccaag 180
258 gcctacgagc gcgaggtgca caacgtgtgg gccacccacg cctgcgtgcc caccgacccc 240
259 aacccccagg agatggtgct ggagaacgtg accgagaact tcaacatgtg gaagaacgac 300
260 atggtggacc agatgcacga ggacatcatc tccctgtggg accagtccct gaagcccctgc 360
261 gtgaagctga cccccctgtg cgtgaccctg aactgcacca cgtgaccaa cgccaccaac 420
262 aacacctaca cggcgagat gaagaactgc tccttcaaca tcaccaccga gctgcgcgac 480
```

RAW SEQUENCE LISTING   DATE: 02/10/2007
PATENT APPLICATION: US/10/572,638   TIME: 09:56:30

Input Set : **N:\DA\pto.da.txt**
Output Set: **N:\CRF4\02102007\J572638.raw**

```
263 aagaagaaga aggagtacgc cctgttctac cgcctggaca tcgtgcccct gaacgagaac 540
264 tcctccgagt accgcctgat caactgcaac acctccgcca tcacccaggc ctgccccaag 600
265 gtgtccttcg accccatccc catccactac tgcgccccccg ccggctacgc catcctgaag 660
266 tgcaacaaca agaccttcaa cggcaccggc ccctgcaaca acgtgtccac cgtgcagtgc 720
267 acccacggca tcaagcccgt ggtgtccacc cagctgctgc tgaacggctc cctggccgag 780
268 gaggagatca tcatccgctc cgagaacctg accgacaacg ccaagaccat catcgtgcag 840
269 ctgaacgagt ccgtggagat cgtgtgcacc cgccccaaca acaacaccg, caagtccatg 900
270 cgcatcggcc ccggccagac cttctacgcc accggcgaca tcatcggcga catccgccag 960
271 gcccactgca acatctccga ggacaagtgg aacaagaccc tgcagcaggt ggccgagaag 1020
272 ctgggcaagc acttccccaa caagaccatc accttcgagc cctcctccgg cggcgacctg 1080
273 gagatcacca cccactcctt caactgccgc ggcgagttct ctactgcaa cacctccaag 1140
274 ctgttcaact ccacctacaa caacaacacc aactccaact ccaccatcac cctgccctgc 1200
275 cgcatcaagc agatcatcaa catgtggcag ggcgtgggcc aggccatgta cgcccccccc 1260
276 atcgccggca acatcacctg caagtccaac atcaccggcc tgctgctgac ccgcgacggc 1320
277 ggcaaggaga acaccaccga gaccttccgc cccggcggcg gcgacatgcg cgacaactgg 1380
278 cgctccgagc tgtacaagta caaggtggtg gagatcaagc ccctgggcgt ggcccccacc 1440
279 gaggccaagc gccgcgtggt ggagcgcgag aagcgcgccg tgggcctggg cgccgtgttc 1500
280 ctgggcttcc tgggcgccgc cggctccacc atgggcgccg cctccatcac cctgaccgtg 1560
281 caggccgcc agctgctgtc cggcatcgtg cagcagcagt ccaacctgct ggcgccatc 1620
282 gaggcccagc agcacatgct gcagctgacc gtgtggggca tcaagcagct gcaggcccgc 1680
283 gtgctggcca tggagcgcta cctgaaggac cagcagctgc tgggcatctg gggctgctcc 1740
284 ggcaagctga tctgcaccac cgccgtgccc tggaactcct cctggtccaa caagtccctg 1800
285 gacgacatct gggacaacat gacctggatg gagtgggacc gcgagatctc caactacacc 1860
286 gacaccatct accgcctgct ggaggagtcc cagaaccagc aggagaagaa cgagcaggac 1920
287 ctgctggccc tggactcctg ggagaacctg tggaactggt tcgacatcac caactggctg 1980
288 tggtacatca agatcttcat catgatcgtg ggcggcctga tcggcctgcg catcatcttc 2040
289 gccgtgctgt ccatcgtgaa ccgcgtgcgc cagggctact ccccccctgtc cttccagacc 2100
290 ctgacccccca acccccgcgg ccccgaccgc ctggagcgca tcgaggagga gggcggcgag 2160
291 caggaccgcg accgctccat ccgcctggtg tccggcttcc tggccctggc ctgggacgac 2220
292 ctgcgctccc tgtgcctgtt ctcctaccac cgcctgcgcg acttcatcct gatcgccgcc 2280
293 cgcaccgtgg agctgctggg ccgctcctcc ctgcgcggcc tgcagcgcgg ctgggaggcc 2340
294 ctgaagtacc tgggctccct ggtgcagtac tggggccagg agctgaagaa gtccgccatc 2400
295 tccctgctgg acaccatcgc catcgccgtg gccgagggca ccgaccgcat catcgaggtg 2460
296 gtgcagcgcg cctgccgcgc catcctgaac atcccccgcc gcatccgcca gggcttcgag 2520
297 gccgccctgc tgtaa                                                 2535
300 <210> SEQ ID NO: 4
301 <211> LENGTH: 2535
302 <212> TYPE: DNA
303 <213> ORGANISM: Human immunodeficiency virus
305 <400> SEQUENCE: 4
306 gccgccatgc gcgtgatggg catcctgcgc aactgccagc agtggtggat ctggggcatc 60
307 ctgggcttct ggatgctgat gatctgcaac gtggtgggca acctgtgggt gaccgtgtac 120
308 tacggcgtgc ccgtgtggaa ggaggccaag accaccctgt tctgcgcctc cgacgccaag 180
309 gcctacgaga aggaggtgca caacgtgtgg gccacccacg cctgcgtgcc caccgacccc 240
310 aacccccagg agatggtgct ggagaacgtg accgagaact tcaacatgtg gaagaacgac 300
311 atggtggacc agatgcacga ggacatcatc tccctgtggg accagtccct gaagccctgc 360
312 gtgaagctga cccccctgtg cgtgaccctg aactgccgca cgtgaccaa cgccaccaac 420
313 aacacctaca cgaggagat caagaactgc tccttcaaca tcaccaccga gctgcgcgac 480
314 aagaagaaga aggtgtacgc cctgttctac cgcctggaca tcgtgccccct gaacgagaac 540
```

VERIFICATION SUMMARY   DATE: 02/10/2007
PATENT APPLICATION: US/10/572,638   TIME: 09:56:31

Input Set : **N:\DA\pto.da.txt**
Output Set: **N:\CRF4\02102007\J572638.raw**

L:10 M:271 C: Current Filing Date differs, Replaced Current Filing Date
L:2428 M:283 W: Missing Blank Line separator, <400> field identifier
L:2429 M:300 W: (50) Intentionally skipped Sequence, : Sequence Id (27) SEQUENCE:
L:14185 M:283 W: Missing Blank Line separator, <400> field identifier
L:14186 M:300 W: (50) Intentionally skipped Sequence, : Sequence Id (149) SEQUENCE: